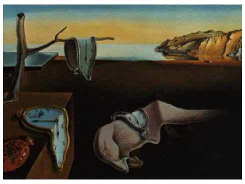

Journal of Statistical Software

April 2009, Volume 30, Book Review 1. *http://www.jstatsoft.org/*

Reviewer: Laurence T. Maloney
New York University

Introduction to Probability with R

Kenneth Baclawski
Chapman & Hall/CRC, Boca Raton, FL, 2008.
ISBN 978-1-4200-6521-3. 363 pp. USD 89.95 (P).
http://www.ccs.neu.edu/home/kenb/stochas/

Introduction to Probability with R by Kenneth Baclawksi is a new addition to Chapman & Hall's series of books featuring R and it is a welcome addition. It is intended to be used by undergraduate students in economics, engineering, computer science, and biology, among other disciplines. It covers core topics in probability theory and its prerequisites and also more advanced topics such as Markov chains. Course slides in Powerpoint and a solutions manual are available to qualified instructors from the publisher.

The book is clearly written and very well-organized and it stems in part from a popular course at MIT taught by the late Gian-Carlo Rota which was originally designed in conjunction with the author of this book. The book goes well beyond the MIT course in making extensive use of computation and R. Teaching simulation methods is of value in itself, of course, but they also allow the student to develop a better understanding of the central results of probability theory. There is a large number of problems some of which are solved in detail. The author also includes historical notes throughout the text giving brief summaries of the history of probability theory and early researchers in probability theory (Pascal, Fermat, Huyghens, . . .). These also serve to capture the students' interest and convey the fascinating history of probability since the 1600s.

While the text makes extensive use of R it does not attempt to teach students how to program. It is likely that a course based on this text would have to be supplemented by a separate text on programming in R and some class time devoted to teaching programming concepts as well as R. This book is a course on probability theory and while a few topics in statistical estimation and hypothesis testing are touched on, this book does not cover introductory statistics. It would serve as an exemplary test for the first semester of a two-semester course on probability and statistics.

The book consists of ten chapters and two appendices. The first chapter develops the set algebra needed to present axiomatic probability theory; this material is hard to present in a way that captures the students' interest. The author complements the presentation of set theory and probability theory with several intriguing problems that go beyond the usual regurgitation of results in the main text. For example, a problem from Chapter 1 (credited to Prendergast):

> Two technicians are discussing the relative merits of two rockets. One rocket has two engines, the other four. The engines used are all identical. To ensure success the engines are somewhat redundant: the rocket will achieve its mission even if half its engines fail. The first technician argues that the four-engine rocket ought to be the better one. The second technician replies, "Although I cannot reveal the failure probability of an engine, because it is classified top secret, I can assure you that either rocket is as likely to succeed as the other."
>
> The first technician replies, "Thank you. What you just told me allows me to compute the failure probability both for an engine and for a rocket."
>
> Can you do this computation also? (Exercise 1.20, p. 21)

Chapter 2 introduces combinatoric problems and R first by using built-in functions to compute factorials and then to study the birthday problem in some detail by writing a function that computes the probability that the birth dates of two or more people in a group of n people have the same month and day. The presentation is clear but compressed (students should have a good idea of what a logarithm is and why it is useful in working with large factorials) and this particular example deserves supplementary treatment in class. The result is a plot of the probability of one or more pairs of coincident birthdays as a function of group size, illustrating how to use R's graphical facilities and also illustrating that probability grows to near-certainty between group sizes of 10 to 40.

Discrete random variables are introduced in Chapter 3 and general random variables in Chapter 4 and the author introduces and makes extensive use of R's built-in functions for each distributional family considered. One slightly unusual and welcome topic in Chapter 4 is consideration of the distribution of order statistics, introduced before discussion of mean, median and mode and their distributions also in Chapter 4. The student who masters these two chapters has learned a useful vocabulary of random variables that could be used in developing simulations in biology, economics, and other areas of the natural and social sciences.

Chapter 5 is devoted to the normal distribution with concomitant introduction of variance and the central limit theorem as well as a rudimentary discussion of hypothesis testing and significance levels. These topics are followed by presentation of the strong law of large numbers and several topics usually reserved to the graduate level: Chebyshev's inequality and Kolmogorov's 0-1 Law. These results are not proven but they are clearly presented and explained. The chapter ends with a discussion of the Cauchy distribution that is remarkably well motivated and – fun! By presenting a distribution for which the law of large numbers does not apply, the author nicely emphasizes the importance of the law of large numbers. The extensive use of R code and graphics allows students to develop real insight into the odd behavior of Cauchy random variables. Where is the weak law of large numbers? In the exercises to the chapter where the student is challenged to prove one version of WLLN from the central limit theorem.

Chapter 6 continues with conditional probability, Bayes' theorem and topics in sampling. Neyman-Pearson hypothesis testing is introduced in one of the problems. Chapter 7 continues with the Poisson distribution and the exponential distribution accompanied by relevant physical examples. One especially useful feature of the chapter is the manner in which the author makes connections between uniform, Bernoulli, and Poison random variables.

The first seven chapters form a remarkably clear presentation of the central topics of probability theory. The remainder of the book consists of advanced topics that make less use of specific features of R and which could be approached in any order. Chapter 8 introduces advanced material that likely would be omitted from most undergraduate courses. These include the Laplace transform and a proof of the central limit theorem as well as consideration of renewal processes and a remarkably concise introduction to Bayes networks. Chapter 9 contains a magnificent summary of entropy and information theory and Chapter 10 concludes the main text of the book with a concise discussion of Markov chains. The two appendices summarize material that complements earlier chapters (Appendix A, random walks and Appendix B, memorylessness).

In summary, *Introduction to Probability with R* is a well-organized course in probability theory. It presupposes that students are comfortable with programming and can pick up R from examples. Alternatively, a second text introducing R could be assigned and covered early in the semester. The text is most impressive in its clear treatment of advanced topics and as such is comparable to *Elementary Probability Theory* by Chung and AitSahlia (2003) or Strang's (1988) treatment of linear algebra. In comparison to Chung's text, *Introduction to Probability in R* will appeal more to sophisticated students with no specific preparation in mathematics beyond calculus but an interest in scientific research and a preference for hands-on investigation of probability theory through simulation in R.

References

Chung KL, AitSahlia F (2003). *Elementary Probability Theory: With Stochastic Processes and an Introduction to Mathematical Finance*. Springer-Verlag, New York, NY.

Strang G (1988). *Linear Algebra and its Applications*. 3rd edition. Brooks Cole, Florence, KY.

Reviewer:

Laurence T. Maloney
New York University
Department of Psychology
Center for Neural Science
6 Washington Place, 2nd Floor
New York, NY 10003, United States of America
E-mail: ltm1@nyu.edu
URL: http://www.psych.nyu.edu/maloney/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 30, Book Review 1 *Published:* 2009-04-27
April 2009